

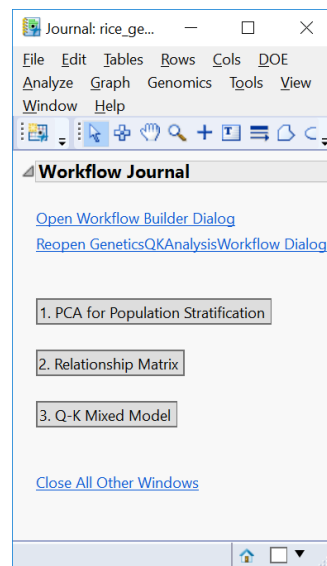
K Matrix Compression

Q-K association analysis is computationally intensive, and it can take a very long time to analyze larger datasets. The part of the analysis that incorporates the K matrix is especially time consuming. The K matrix is square, so for every individual or line in the study, a corresponding K matrix variable must be added. In our example with 108 lines, there are 108 K matrix variables included in each model.

There is a technique for reducing the number of variables required to represent the familial relatedness between lines. With fewer variables each model, run time is significantly reduced. The technique is called **K Matrix Compression** (Zhang et al., 2010). It can be performed in JMP Genomics as part of the **Genetics Q-K Analysis Workflow**, or as a free-standing process. The algorithm optimizes the compression for one trait variable at a time, so it needs to be repeated for each trait to be analyzed.

Let's work through an example using the **Genetics Q-K Analysis Workflow** (See *QK Analysis Workflow Step-Guide*). We will add the **K matrix compression** to the previous analysis.

1. Find the journal output from the previous **Genetics Q-K Analysis Workflow**. Click the button labeled **Reopen GeneticsQKAnalysisWorkflow Dialog**.



- The dialog for the workflow reopens, with the previous settings loaded.
2. On the **General** tab, remove all the other traits from the **Trait Variables** box, leaving only **FL**.
 - Note: The K matrix can be compressed for only a single trait at a time.
 3. Assign a new **Output Folder** to avoid overwriting the previous results.
 4. On the **K Matrix Options** tab, check the box labeled **Compress the K Matrix**.

5. In the **Compression Rate** box, type “0.1” (without the quotes).
 - This will cause the compression algorithm to evaluate solutions starting with all 343 variables, reducing the number by 10 percent each iteration.
6. Move the variable **GID** to the **ID Variables** box.
7. Click **Run** to start the analysis.
8. When the results journal appears, click on the **Results** link in the **K Matrix Compression** section.
 - The plot shows the evaluation of different amounts of compression. The full, uncompressed K matrix is at the right side of the graph, with all 343 variables, and the amount of compression increases to the left.
 - The three criteria that are graphed have low values for better-fitting models. The lowest value for all three criteria occurs at just over 300 clusters, but there is another similar value at 225. Let’s return to the dialog and force the algorithm to select 225 clusters.
9. Click the button labeled **Reopen Dialog**.
 - The dialog for the **K Matrix Compression** process opens with the previous options loaded. Note that this is not the original workflow dialog, it is the standalone dialog for **K Matrix Compression**. The workflow shows a simplified interface for the standalone tool, now we are using the more complex and robust interface of the standalone tool.
10. In the **Analysis** tab, change the **Compression Method** to **Automated**.
11. Enter “225” (without the quotation marks) for the **Number of Clusters for Automated Compression**.
12. Click **Run** to start the analysis.
13. When the results appear, click the **Q-K Mixed Model** button.
 - We are operating outside the workflow now, so we must launch each process individually.
14. Fill out the **Annotation** tab
15. Click **Run** to start the analysis.
 - The results appear similar to the uncompressed analysis.

Summary

Compression of the K-Matrix can be used to reduce computing time in **Q-K Association Analysis** while still producing similar results to analysis with an uncompressed matrix.